Sleep homeostasis and models of sleep regulation

Achermann, Peter ; Borbély, Alexander A

Sleep Homeostasis and Models of Sleep Regulation

Peter Achermann; Alexander A. Borbély

Chapter 36

Chapter Highlights

- Sleep homeostasis refers to the aspect of sleep regulation dependent on sleep and waking, as homeostatic mechanisms counteract deviations from an average reference level of sleep. These mechanisms augment sleep propensity when sleep is curtailed or absent, and they reduce sleep propensity in response to excess sleep. In general, homeostasis can be defined as "the coordinated physiological processes, which maintain most of the steady states in the organism."[1,2]
- Non–rapid eye movement sleep is not a homogeneous substate of sleep but can be subdivided according to the predominance of slow waves in the electroencephalogram (EEG). One of the most important functional EEG parameters is referred to as slow wave activity; it is equivalent to delta activity and encompasses components of the EEG signal in the frequency range of approximately 0.5 to 4.5 Hz as obtained by spectral analysis. Under physiologic conditions, this EEG variable can be regarded as an indicator of sleep depth or sleep intensity.[7]
- The two-process model postulates that a homeostatic process rises during waking and declines during sleep and interacts with a circadian process that is not directly dependent of sleep and waking. The model served as a conceptual framework and inspired many experiments to test its predictions.[3,27,83]
- Models may address processes at different levels (from the microscopic [cellular] level to the macroscopic [systemic] level) and at different time scales (from the range of milliseconds or seconds up to hours or days).[74] Many mathematical models of sleep regulation represent extended versions of the two-process model incorporating neurophysiologic processes. Being aware of the power and limitations of models is important for selecting the most appropriate one for the question to be addressed.

Three distinct processes underlie sleep regulation. A homeostatic process, whose level is a function of prior sleep and waking, plays a major role in sleep regulation. Sleep is also modulated by a circadian process, a clocklike mechanism that is independent of prior sleep and waking. An ultradian process occurs within sleep and is represented by the alternation of the two basic sleep states: non–rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep.

One of the main topics of this chapter is sleep homeostasis. *Homeostasis* has been defined as "the coordinated physiological processes, which maintain most of the steady states in the organism."[1] The term *sleep homeostasis*[2] refers to the aspect of sleep regulation dependent on sleep and waking, as homeostatic mechanisms counteract deviations from an average reference level of sleep. These mechanisms augment sleep propensity when sleep is curtailed or absent, and they reduce sleep propensity in response to excess sleep.

The interest in the modeling approach to sleep regulation has increased over the past decades. Models help delineate the processes involved in the regulation of sleep and thereby offer a conceptual framework for the analysis of existing and new data.[3] For reviews of modeling circadian rhythms related to sleep, see Roenneberg and colleagues,[4] Beersma,[5] and Klerman and St Hilaire.[6]

HOMEOSTATIC REGULATION OF SLEEP

Electroencephalographic Slow Wave Activity: A Physiologic Indicator of NREM Sleep Homeostasis

Slow Wave Sleep and Slow Wave Activity

NREM sleep is not a homogeneous substate of sleep but can be subdivided according to the predominance of slow waves in the electroencephalogram (EEG). One of the most important functional EEG parameters is referred to as slow wave activity (SWA). It is equivalent to delta activity and encompasses components of the EEG signal in the frequency range of approximately 0.5 to 4.5 Hz as obtained by spectral analysis.[7]

In addition to delta waves, a low-frequency component with a mean peak value of 0.7 to 0.8 Hz is present in the EEG power spectrum of NREM sleep.[8–10] The typical decline in delta activity from the first to the second NREM sleep episode was not present at frequencies below 2 Hz.[8] This could indicate that the low-frequency component reflects a separate process that differs from the one represented by delta activity. Alternatively, the results could be due to a frequency shift in the course of sleep.[11] The changes of the low-frequency component in response to naps did not differ from those of SWA.[12] Increased sleep pressure resulted in a redistribution of waves

between 0.5 and 2 Hz: The number of waves per minute was reduced below 0.9 Hz but was increased above 1.2 Hz; EEG power was increased only above 1 Hz.[13] Periodicities at even lower frequencies include the recurrence of sleep spindles at 4-second intervals,[8,14] the tendency of slow waves to recur at 20- to 30-second intervals,[8] and a cortical infra slow oscillation (0.02 to 0.2 Hz).[15,16]

Slow Waves and Sleep Intensity

It was recognized as early as 1937 that sleep intensity is reflected by the predominance of slow waves in the sleep EEG.[17] Subsequent studies confirmed that the responsiveness to stimuli decreases as EEG slow waves become more predominant.[18] Under physiologic conditions, this EEG variable can be regarded therefore as an indicator of sleep depth or sleep intensity.

Global Time Course of Slow Wave Activity during Sleep

Slow wave sleep (SWS; the high-intensity part of NREM sleep) appeared to be a good candidate for a physiologic marker of sleep homeostasis. The predominance of SWS in the early part of the sleep episode was documented in several early studies.[18-20] All-night spectral analysis of the sleep EEG made it possible to quantify SWA and to delineate its time course during sleep.[7] Its mean value per cycle plotted for consecutive NREM-REM sleep cycles showed a monotonic decline over the first three cycles. Figure 36-1 (*left*) shows the changes of mean EEG power density over four cycles for the frequency range between 0.25 and 20 Hz. The values of each bin are expressed relative to the reference level of cycle 4 (100%). Note that although the largest changes occur in the low delta range, they encompass frequencies up to the theta band.

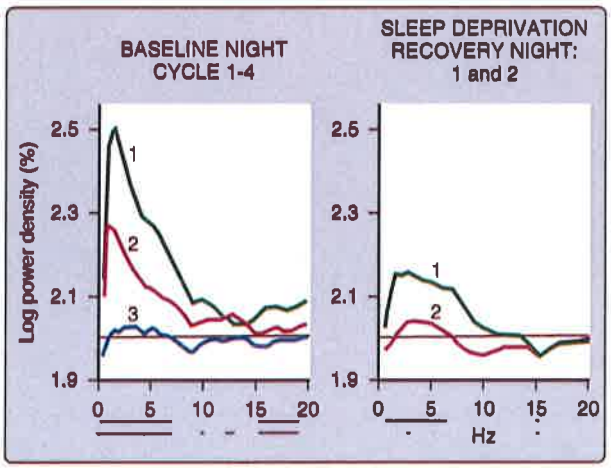

Figure 36-1 *Left*, Changes of relative spectral power density over the first four NREM-REM sleep cycles of a baseline night (N = 8). In each frequency bin the data are expressed relative to the value in the fourth cycle (100%; *horizontal line*). *Right*, Effect of sleep deprivation (40.5 hours waking) on spectra of the sleep electroencephalogram (EEG). In each bin, the values of the first two recovery nights are plotted relative to the baseline night (100%). The *upper* and *lower horizontal bars* below the abscissa indicate for the left part significant differences between cycles 1 and 2, and between cycles 2 and 3, respectively, and for the right part between recovery 1 and baseline, and between recovery 2 and baseline, respectively. (Modified from Borbély AA, Baumann F, Brandeis D, et al. Sleep deprivation: effect on sleep stages and EEG power density in man. *Electroencephalogr Clin Neurophysiol* 1981;51:483–93.)

Nap Studies

The analysis of daytime naps is useful for assessing the level of SWA after various durations of waking. Naps taken later in the day contained more SWS than naps taken earlier in the day. In a detailed study of daytime naps scheduled at 2-hour intervals throughout the day, direct evidence for a monotonic rise of SWA was obtained.[12,21-23] If naps reverse the rising trend of slow wave propensity, a reduction of SWA in the subsequent nighttime sleep can be expected. This prediction was borne out by the results of several experiments (see Werth et al.[24] for literature references). Furthermore, when the duration of nighttime sleep was shortened, SWA in a subsequent morning nap was enhanced.[25,26]

Effect of Sleep Deprivation

It has repeatedly been shown that partial or total sleep deprivation gives rise to increased SWS in the recovery night (see Borbély[27] for a review of the older literature). Webb and Agnew[20] presented compelling evidence that SWS increases as a function of prior wake duration. The quantitative assessment of SWA using all-night spectral analysis revealed that a night without sleep (i.e., 40.5 hours of wakefulness) resulted in an enhancement of this EEG variable during recovery sleep.[7] Figure 36-1 (*right*) illustrates the changes of power density in the two recovery nights relative to the baseline level (100%). In the first recovery night, the largest increase was present in the low delta range, the part of the spectrum undergoing the largest changes in the course of baseline sleep (see Figure 36-1, *left*).

Figure 36-2 depicts the global trend as well as the ultradian dynamics of SWA over successive NREM-REM sleep cycles. The prolongation of the waking period causes a prominent rise of SWA during recovery sleep. A declining trend over three to four cycles is evident in both records. Note that the peaks are at a steady low level during the last four cycles of recovery sleep.

The enhancement of SWA by sleep deprivation was confirmed in numerous studies[28,29] (for references before 1992, see Borbély and Achermann[30]). The extent of the increase was shown to be a function of the duration of prior waking.[21,31]

Selective Slow Wave Deprivation

The propensity of SWA does not necessarily dissipate during sleep but may persist at an elevated level if SWS is prevented. Thus suppression of slow waves by acoustic stimuli during the first 3 hours of sleep resulted in a prominent rise of SWA after the stimuli were discontinued.[32] In another study, daytime sleep episodes with and without SWS deprivation were compared.[33] The experimental suppression of SWS during an interval corresponding to 90% of the undisturbed episode resulted in an increased accumulation of SWS and an extension of sleep duration. Topographic studies revealed regional differences in the effectiveness of and response to selective SWS deprivation.[34] Taken together the results indicate that slow waves are not merely an epiphenomenon of sleep but instead reflect major sleep-regulating mechanisms. This is supported by the findings that selective slow wave deprivation impairs perceptual[35] and visuomotor learning.[36] In another study, two nights of SWS disruption increased sleepiness but had only minor effects on daytime functioning.[37]

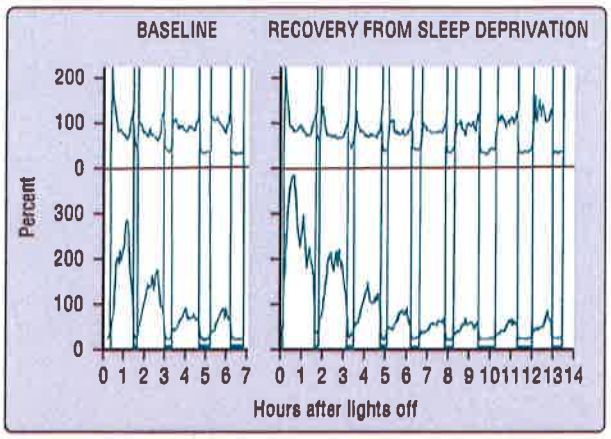

Figure 36-2 Time course of slow wave activity (power in the 0.75- to 4.5-Hz band; *lower curves*) and activity in the spindle frequency range (13.25- to 15.0-Hz; *upper curves*) recorded under baseline conditions and after sleep deprivation (36 hours of wakefulness). NREM sleep episodes were subdivided into 20 equal intervals, and REM sleep episodes were divided into 5 intervals. Mean values per interval were calculated before averaging across subjects (N = 8 except for cycle 8 of recovery sleep where N = 6) and were expressed relative to the mean level in baseline NREM sleep (100%). The mean timing of REM sleep episodes is delimited by *vertical lines* and *horizontal bars* above the abscissa. (Reanalysis of the data from Dijk DJ, Brunner DP, Borbély AA. Time course of EEG power density during long sleep in humans. *Am J Physiol* 1990;258:R650–1, by D. Aeschbach.)

Ultradian Dynamics of Slow Wave Activity and Spindle Frequency Activity

Buildup of Slow Wave Activity within NREM Sleep Episodes

The mean level and the peak of SWA are determined not only by the duration of prior waking and sleep but also by the rise rate within single NREM sleep episodes.[38-40] It is evident from Figure 36-2 that the rise rate of SWA decreases over the first three episodes both under baseline conditions and during recovery from sleep deprivation. In addition, the effect of prolonged waking manifests itself in a steeper buildup (of SWA) within NREM sleep episodes.[28,29,40] A decrease in the slope of single slow waves was associated with a reduction of sleep pressure during sleep in humans and animals,[41,42] a change that could be simulated in a model by reducing synaptic strength.[43] However, amplitude and slope of slow waves were highly correlated in the 0.5- to 2-Hz range.[13]

Slow Wave Activity and Spindle Frequency Activity

The term *spindle frequency activity* (SFA) is used to denote the power in the frequency range of sleep spindles (12 to 15 Hz). There is a close correspondence between this measure and measures based on the occurrence of sleep spindles.[28]

The time courses of SWA and SFA differ in several respects. The global decline of SWA does not occur in the spindle frequency range. Within NREM sleep episodes, SFA shows a bimodal pattern with an initial and a terminal peak. This gives rise to a U-shaped curve within the episode and a partly inverse relationship to SWA[28,44-49] (see Figure 36-2).

Regulation of REM Sleep

The principles underlying REM sleep regulation seem to be more complex than those for NREM sleep regulation. This was evident in a selective REM sleep deprivation experiment,[50] in which two salient observations were made that were difficult to reconcile. On the one hand, REM sleep deprivation necessitated a dramatic rise in the frequency of interventions during the night to prevent this sleep state. On the other hand, there was a modest rise in the number of interventions across the three consecutive deprivation nights, and the 40% REM sleep rebound in the first recovery night by no means compensated for the amount of REM sleep lost. Two hypotheses were advanced. In the first, it is assumed that the homeostatic drive is strong, which is reflected by the dramatic rise in interventions during the deprivation nights. Waking may in part substitute for REM sleep, thereby accounting for the moderate night-to-night increase in interventions and the small REM sleep rebound. According to the second hypothesis, the homeostatic drive for REM sleep is weak, and the rising trend in the number of interventions is attributed to circadian factors as well as to a sleep-dependent disinhibition of REM sleep propensity. This hypothesis could explain the limited savings from one night to the other as well as the modest rebound. For a recent review on possible functions of REM sleep, see Vyazovskiy and Delogu.[51]

NREM versus REM Sleep Homeostasis

Effect of NREM Sleep Pressure on REM Sleep Homeostasis

During recovery from total sleep deprivation, SWS and EEG SWA exhibit an immediate rebound, whereas the increase in REM sleep is delayed to subsequent nights or is not present at all. Selective REM sleep deprivation augments REM sleep pressure, which is manifested by the increasing number of interventions required to prevent REM sleep episodes (for the older literature, see Borbély[27]). However, the occurrence of REM sleep rebound during recovery sleep is smaller than expected on the basis of the deficit.[50,52] This suggests that REM sleep is not as finely regulated as NREM sleep. However, this notion is contradicted by partial sleep deprivation studies. A REM sleep deprivation in the first 5 hours of sleep induced a REM sleep rebound in the subsequent 2.25 hours.[53] A curtailment of sleep duration during two or four nights, which induced a substantial REM sleep deficit, was followed by REM sleep rebound in the two recovery nights.[54,55] In these experiments, the REM sleep rebound occurred at a time when slow wave pressure was either low at the end of sleep[53] or was much less increased than REM sleep pressure.[54,55] These results also suggest that REM sleep is indeed regulated but that the manifestation of REM sleep homeostasis is hampered by an elevated slow wave pressure.

Effect of REM Sleep Pressure on the NREM Sleep Electroencephalogram

In accordance with the notion of a mutual inhibitory interaction of the factors controlling SWA and REM sleep,[27] not only is REM sleep inhibited by slow wave pressure, but SWA is also inhibited by REM sleep pressure. Thus selective REM sleep deprivation led to a significant reduction in the low-frequency activity of the NREM sleep EEG,[53] an observation that was also made in an animal experiment.[56] The rise in REM sleep pressure induced by repeated partial sleep deprivation suppressed the typical low-delta peak in the NREM sleep spectrum.[54,55] However, this effect was not seen after selective REM sleep deprivation.[50]

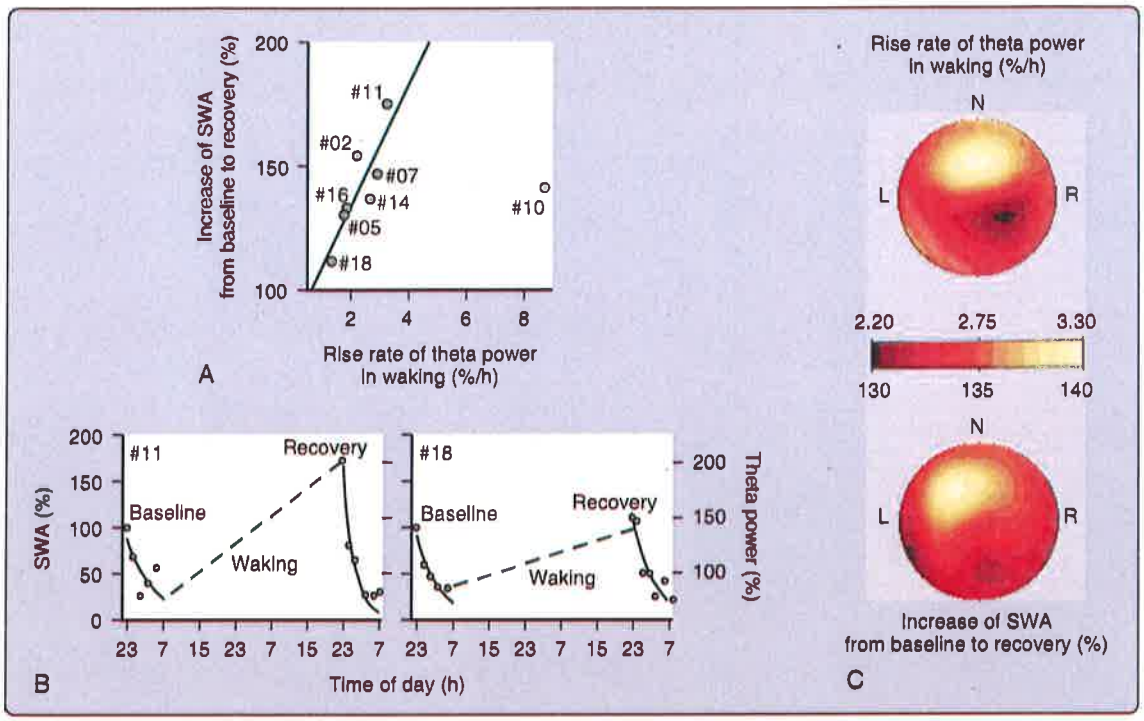

Figure 36-3 **A,** Relationship between homeostatic markers of the sleep electroencephalogram (EEG) and waking EEG. Increase (%) of slow wave activity (SWA; 0.75–4.5 Hz) in the first NREM sleep episode from baseline to recovery sleep is plotted as a function of the rise rate (%/h) of theta power (5.0–8.0 Hz) in waking. The linear regression line fitted through 7 data points is indicated ($r = 0.851$, $r^2 = 0.724$, $P = .015$). Subject no. 10 was excluded from the regression. **B,** Association between rise of SWA in sleep and theta activity in waking illustrated for two subjects. Mean SWA per NREM sleep episode is plotted at the beginning of each episode and expressed relative to the baseline value of the first NREM sleep episode (100%). Exponential functions were fitted through the data points (*solid curves*). The regression line represents theta power in waking (*interrupted line*). **C,** Topographic distribution of the rise rate of theta power (*top*) during waking and of the increase of SWA (*bottom*) in the first NREM sleep episode from baseline to recovery sleep. Maps are based on 27 EEG derivations (average reference; extended 10-20 system). Values are plotted on a color scale at the corresponding position on the planar projection of the hemispheric scalp model. Values between electrodes were linearly interpolated. (Modified from Finelli LA, Baumann H, Borbély AA, Achermann P. Dual electroencephalogram markers of human sleep homeostasis: correlation between theta activity in waking and slow-wave activity in sleep. *Neuroscience* 2000;101:523–9.)

Homeostatic Marker In the Waking Electroencephalogram

It had been shown in early studies that power in the theta band (theta activity) and alpha activity of the waking EEG is associated with sleepiness[57,58] and that total[58] or partial sleep deprivation[55] enhanced the power in these frequency bands. A saturating exponential function with a time constant of 18.18 hours was reported to fit the rise of theta activity in the waking EEG.[59] Spectral analysis showed that the largest changes occurred in the theta band (see Borbély and Achermann[60] for review). These undergo a circadian modulation in addition to the changes related to wake time.[61–66] During prolonged wakefulness, subjective sleepiness correlated positively with theta activity with a focus in frontal derivations and negatively with alpha activity at all derivations.[67] A forced desynchrony study with a scheduled waking episode of 28 hours showed a monotonic rise of delta and beta activity in the frontocentral derivation.[65] An analysis of persons subjected to sleep deprivation revealed that the rise rate of theta activity in the waking EEG is correlated with the increase of SWA in the first NREM sleep episode of recovery sleep (Figure 36-3, *A* and *B*).[61] Moreover, both effects were largest in frontal areas (Figure 36-3, *C*). In summary, theta activity in waking and SWA in sleep may be markers of a common homeostatic sleep process.

Independence and Interactions of Homeostatic and Circadian Processes

There is evidence that homeostatic and circadian facets of sleep regulation can be independently manipulated and therefore may be controlled by separate mechanisms. Thus throughout a 72-hour sleep deprivation period, the subjective alertness ratings continued to show a prominent circadian rhythm.[68] Conversely, in a study in which the phase of the circadian process (as indexed by body temperature and plasma melatonin) was shifted by bright light in the morning, the time course of SWA remained unaffected.[69]

A powerful experimental paradigm is the forced desynchrony schedule in which the homeostatic and circadian facet of sleep can be separately analyzed.[70,71] In this long-term protocol, the imposed sleep-wake cycle (e.g., 20 hours or 28 hours) lies outside the range of entrainment of the circadian pacemaker. When this schedule is maintained for 3 weeks,

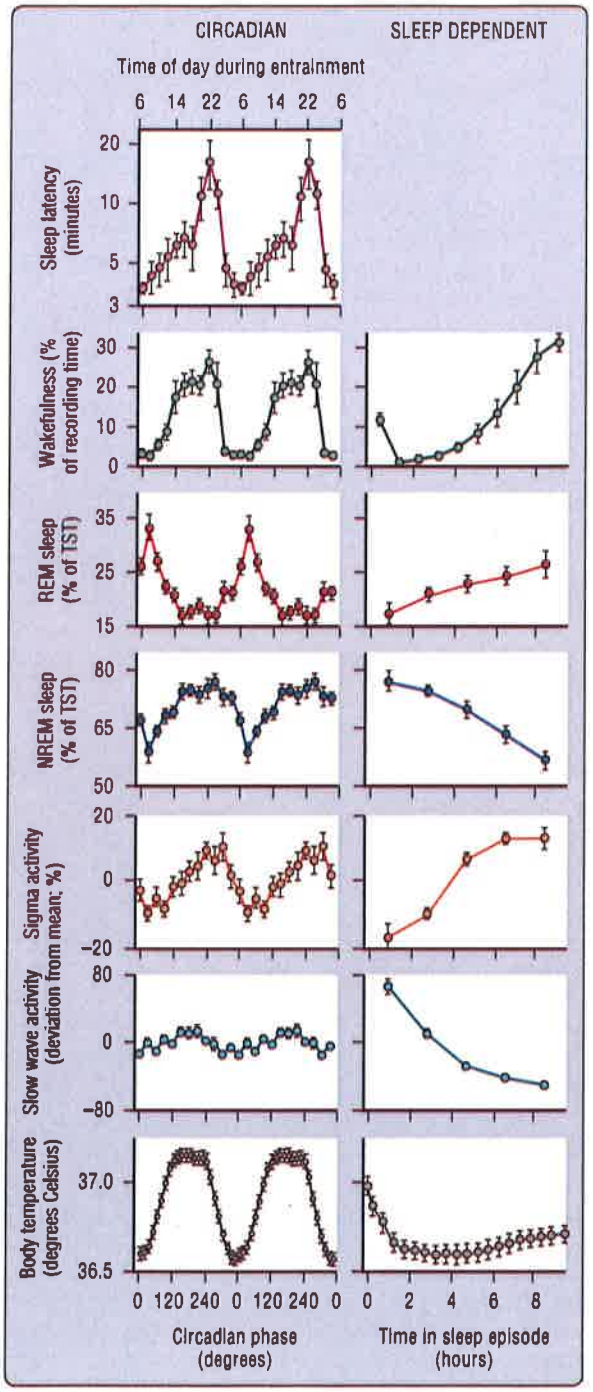

Figure 36-4 Circadian and sleep-dependent or homeostatic factors in sleep regulation. The main effects of circadian phase and sleep homeostasis on sleep were analyzed by aligning the data relative to the circadian component of the body temperature cycle (*left panels*) or the beginning of the sleep opportunity (*right panels*). Slow wave activity shows weak circadian and strong sleep-dependent modulation; sigma (sleep spindle) activity shows strong circadian and sleep-dependent modulation. NREM sleep percentage shows equal circadian and sleep-dependent components. REM sleep percentage shows marked circadian maximum just after the temperature nadir and sleep-dependent increase (disinhibition). Wakefulness in scheduled sleep episodes shows that the circadian drive for wakefulness is maximum 7 to 9 hours before temperature nadir, which is 1 to 3 hours before habitual bedtime; there is a strong wake-dependent increase. Sleep latency shows strong circadian modulation; the longest sleep latencies occur 7 to 9 hours before the body temperature nadir, and the shortest sleep latencies occur at the body temperature nadir. TST, Total sleep time. (Modified from Dijk DJ, Czeisler CA. Contribution of the circadian pacemaker and the sleep homeostat to sleep propensity, sleep structure, electroencephalographic slow waves and sleep spindle activity in humans. *J Neurosci* 1995;15:3526–38; and Dijk DJ, Czeisler CA. Paradoxical timing of the circadian rhythm of sleep propensity serves to consolidate sleep and wakefulness in humans. *Neurosci Lett* 1994;166:63–8. Corresponds to Figure 34-4 of Dijk DJ, Franken P. Interaction of sleep homeostasis and circadian rhythmicity: dependent or independent systems? In: Kryger MH, Roth T, Dement WC, editors. *Principles and practice of sleep medicine*. 4th ed. Philadelphia: Saunders; 2005. p. 418–34.)

sleep occurs at different circadian phases. The contributions of the homeostatic and circadian components can be estimated by folding the data of the variable investigated either at the period of the imposed sleep-wake cycle or at the period of circadian rhythm. Various claims of the two-process model were supported by the results obtained in this paradigm.[70] As shown in Figure 36-4, the variation of SWA is accounted for mainly by homeostatic (i.e., sleep-wake dependent) factors, whereas the percentage of REM sleep, NREM sleep, SFA (sigma activity), and sleep consolidation are determined by both homeostatic and circadian factors. Furthermore, a previously postulated sleep-related disinhibition of REM sleep[27] was confirmed. Severe sleep restriction attenuates the circadian modulation of sleep.[71,72] The role of the metrics used and whether homeostatic and circadian processes exhibit a linear or a nonlinear interaction are still under discussion.[73]

MODELS OF SLEEP REGULATION

Models help delineate the processes involved in the regulation of sleep and thereby offer a conceptual framework for the analysis of existing and new data. In addition, they inspire new experiments to test predictions of the models.

Models may address processes at different levels (from the microscopic [cellular] level to the macroscopic [systemic] level) and at different time scales (from the range of milliseconds or seconds up to hours or days[74]). Being aware of the power and limitations of models is important for selecting the most appropriate one for the question to be addressed.

The term *model* has been loosely applied to hypothetical descriptions of events, features, and processes related to sleep. The following synopsis is restricted to major models that include a mathematical description and address sleep regulation. Their main features are summarized in Tables 36-1 and 36-2. We describe only selectively models addressing the generation of specific features of the sleep EEG and do not include molecular state variables of sleep regulation.[75]

Table 36-1 Two-Process Model and Related Models

Designation	Assumption	Description and Comment
Two-process model[27,78,82,83]	Sleep propensity is determined by a homeostatic process S and circadian process C. The interaction of S and C determines the timing of sleep and waking.	Time course of S derived from EEG slow wave activity; phase position and shape (skewed sine wave) of C derived from sleep duration data obtained at various times of the 24-hr cycle.
Model of ultradian variation of slow wave activity[84-86]	Derived from the two-process model. The level of S determines the buildup rate and the saturation level of slow wave activity within NREM sleep episodes.	In contrast to the original two-process model, the change of S, not the level of S, corresponds to slow wave activity; that is, the decline of S is proportional to the amount of slow wave activity. A REM sleep oscillator triggers the decline of slow wave activity before REM sleep.
Three-process model of sleepiness-alertness regulation[97-99,167]	Sleepiness and alertness are simulated by the combined action of a homeostatic process, a circadian process, and sleep inertia (process W). Extension to include performance, sleep latency, and sleep length. Adaptation of homeostatic process to account for performance prediction in sleep-restriction protocols.	Parameters derived from rated sleepiness during sleep-wake manipulations. Alertness nomogram for sleep-related safety risks.
Interactive mathematical models of alertness and cognitive throughput[100]	Alertness and cognitive throughput are determined by a nonlinear interaction of a homeostatic (H) and a circadian (C) process. Sleep inertia is also included. H shows a sigmoidal decline during waking and a saturating exponential increase during sleep at a rate determined by the circadian phase.	Parameters derived from sleep inertia studies, sleep deprivation studies initiated across all circadian phases, and 28-hr forced desynchrony studies.

Two-Process Model and Related Models

The two-process model and related models are summarized in Table 36-1. Further models inspired by the two-process model addressing performance simulations were discussed in a special issue of *Aviation, Space and Environmental Medicine*[76] and elsewhere.[3,30,60,77-79]

The relationship between SWS and the duration of prior waking has been documented by Webb and Agnew[20] and placed into a theoretical framework by Feinberg.[80] The two-process model, originally proposed to account for sleep regulation in the rat,[2,81] postulates that a homeostatic process (process S) rises during waking and declines during sleep and interacts with a circadian process (process C) that is not directly dependent on sleep and waking. The time course of the homeostatic variable S was derived from EEG SWA (Figure 36-5, *A*). Various aspects of human sleep regulation were addressed in a qualitative version of the two-process model.[27] An elaborated, quantitative version of the model was established in which process S varied between an upper and a lower threshold that are both modulated by a single circadian process.[82,83] This model was able to account for such diverse phenomena as recovery from sleep deprivation, circadian phase dependence of sleep duration, sleep during shift work, sleep fragmentation during continuous bed rest, and internal desynchronization in the absence of time cues.[83]

In a later version of the model (proposed by Beersma et al.[21] and Dijk et al.[32] and formalized by Achermann and Borbély[84,85]), it is the change of S, and not its level, that is proportional to the momentary amount of SWA. The elaborated model addressed not only the global changes of SWA as represented by process S but also the changes within NREM sleep episodes. In general, a close fit was obtained between the simulated and empirical SWA data and their time course (Figure 36-5, *D*). In particular, the occurrence of late SWA peaks during extended sleep could be simulated (Figure 36-5, *C*). The simulations demonstrated that the model accounts in quantitative terms for empirical data and predicts the changes induced by the prolongation of waking or sleep. This version of the model was used to simulate the dynamics of SWA in an experimental protocol with an early evening nap[24] and the effect of changes in REM sleep latency on the time course of SWA.[86] Zavada and colleagues[87] used this model's approach to investigate regional aspects of sleep homeostasis.

The homeostatic process S is modeled by two exponential functions, one describing the rising limb during waking, the other the declining limb during sleep (see Figure 36-5, *A*). The buildup of sleep pressure is approximated by a saturating exponential function (time constant of increase; upper asymptote), its dissipation by a declining exponential function (time constant of decrease; lower asymptote). Thus, process S oscillates between an upper and lower asymptote (relaxation oscillator), its dynamics being determined by the distance between the asymptotes and by the time constants. The time constants show significant interindividual variability[88] and vary also across brain regions.[89] Importantly, homeostasis is mainly reflected in the time constants, and it was hypothesized that a slower buildup may be associated with an increased tolerance to sleep deprivation.[88] A slowing of the buildup of sleep pressure was observed in preschool children in the course of development[90] and during early adolescence.[91] It is important to recognize that the absolute values of SWA are highly variable

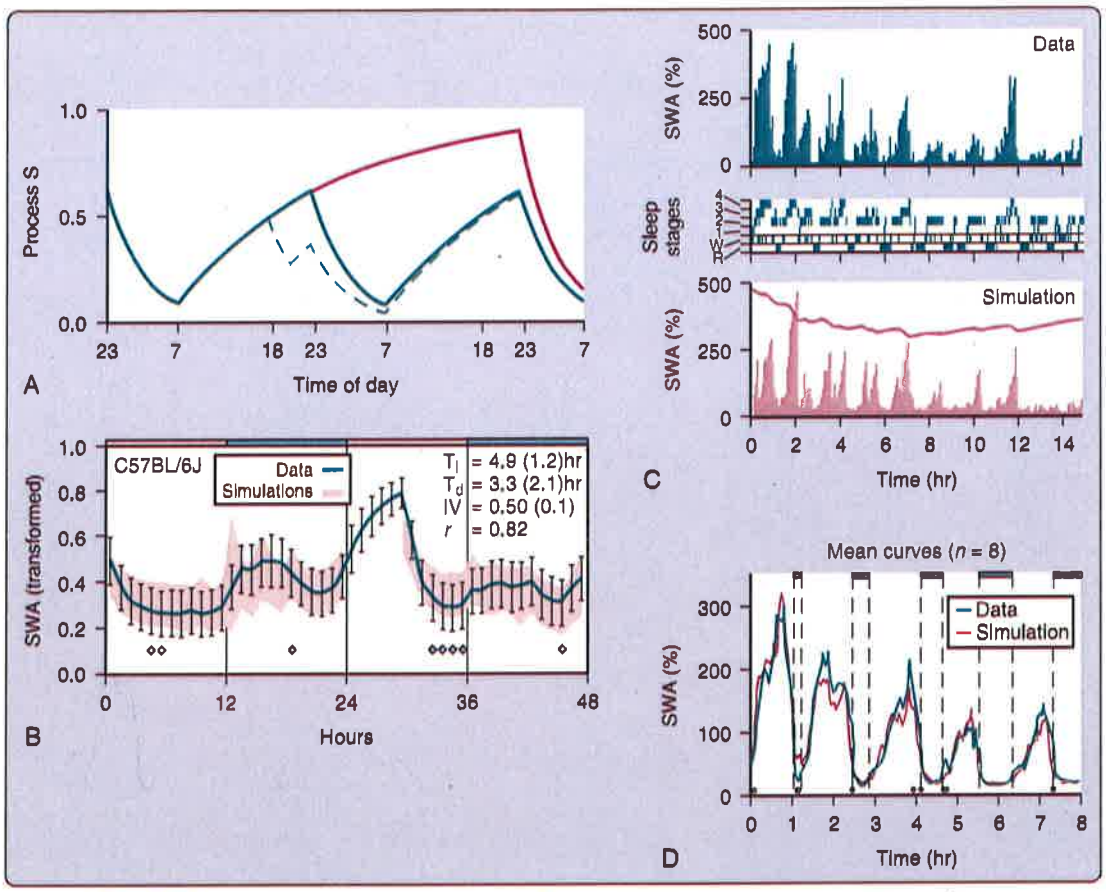

Figure 36-5 Two-process model of sleep regulation. **A,** Simulations of the homeostatic process S according to different experimental conditions. The *green line* indicates baseline condition with an 8-hour sleep episode; the *red line* indicates sleep deprivation (40 hours of wakefulness) and recovery sleep; the *dashed line* indicates effect of a 2-hour daytime nap at 18:00 hours. **B,** Sleep regulation in the mouse. Time course of slow wave activity (SWA) and simulation with the optimized time constants for the increase (Ti) and decrease (Td) and initial value (IV) of process S for C57BL/6J mice (N = 8). *Curves* and *shaded areas* connect 1-hour mean values (±SEM) for 24-hour baseline, 6-hour sleep deprivation, and 18-hour recovery. The close fit between the simulation of process S (*pink areas*) and time course of empirical SWA (*solid line*) indicates that the model can predict SWA from the temporal organization of sleep. *Diamonds* indicate differences between simulation and data ($P < .05$; two-tailed paired t-test). For the comparison between SWA and S, SWA was transformed according to a linear regression. *Inset:* Mean values of Ti, Td, and IV (SEM) and the mean *r* value of the fit between SWA and S. **C,** Empiric SWA (*top*), sleep stages (*center*), and simulation of SWA and process S (*bottom*) of an individual extended baseline sleep episode starting at 00:00 hours (prior waking: 17:00 hr). Empiric and simulated SWA were standardized with respect to the mean value of the first 7 hours of sleep. Values are plotted for 1-minute intervals. **D,** Mean empirical (*green line*) and simulated SWA (*pink line*) (N = 8) of an extended baseline experiment (analysis of first 8 hours). Significant differences are indicated by *open circles* (paired t-test; $P < .05$). *Bars on top* and the *interrupted vertical lines* indicate REM sleep episodes (mean values). (**B** modified from Huber R, Deboer T, Tobler I. Effects of sleep deprivation on sleep and sleep EEG in three mouse strains: empirical data and simulations. *Brain Res* 2000;857:8–19. **C** and **D** modified from Achermann P, Dijk DJ, Brunner DP, Borbély AA. A model of human sleep homeostasis based on EEG slow-wave activity: quantitative comparison of data and simulations. *Brain Res Bull* 1993;31:97–113.)

between individuals and are in a large part dependent on age.[92] Thus absolute levels of SWA are not a measure of sleep pressure or homeostasis per se, but it is the relative change in SWA in response to a challenge that is informative (e.g., for total or partial sleep deprivation, sleep restriction, naps). Furthermore, the distance between the asymptotes may be interpreted as the capacity of the brain to generate slow waves.[88,90,91]

Finally, the data analysis showed that not only the timing of sleep but also the changes in daytime vigilance are governed by the interaction of processes S and C, as simulated by Daan and colleagues.[83] The rising homeostatic sleep pressure during waking seems to be compensated by the declining circadian sleep propensity.[83,93–95] Conversely, during sleep the rising circadian sleep propensity might serve to counteract the declining homeostatic sleep pressure, thereby ensuring the maintenance of sleep.[96]

Based on a similar concept, the changes of subjective sleepiness/alertness ratings were simulated by a combined action of a homeostatic process (S), a circadian process (C), and a process representing sleep inertia (W) (three-process

Table 36-2 Neurophysiologic Network Models of Sleep-Wake Regulation (Brainstem and Hypothalamic Control of Sleep-Wake States)

Designation/Model	Assumption*	Description/Comment
Quantitative model of sleep-wake dynamics based on physiology of the brainstem ascending arousal system Phillips and Robinson[125,132-137,168-171]	Model includes VLPO (where circadian and homeostatic drives enter the system), monoaminergic and cholinergic nuclei of the ascending arousal system, and corresponding interconnections Mutual inhibition between wake-promoting monoaminergic group and sleep-promoting VLPO causes flip-flop behavior REM sleep not incorporated	Human sleep: simulation of basic sleep behavior; effects of sleep deprivation, effect of caffeine, fatigue, impulsive stimuli, internal desynchronization, and shift work Animal sleep: simulation of interspecies differences and unihemispheric sleep
Biologically based mathematical model of the sleep-wake cycle Rempe et al.[126]	Model is based on flip-flop conceptual models for sleep-wake and REM-NREM sleep alterations Includes the sleep-promoting neurons in the VLPO, the wake-promoting monoaminergic cell groups, and orexin neurons Mutual inhibition of REM-on and REM-off populations	Human sleep: simulation of basic sleep patterns; sleep deprivation; role of orexin; and sleep-onset REM sleep
Mathematical model of neuronal regulation of waking, NREM sleep, and REM sleep Kumar et al.[172]	Two distinct flip-flop circuits, one governing wake-NREM sleep transitions (POAH, MRF, CRF, and orexinergic groups) and the other NREM sleep–REM sleep transitions (LDT/PPT and LC) REM sleep generation by presynaptic inhibition of substantia nigra onto REM-off terminals projecting on REM-on neurons. Putative REM sleep homeostasis. Mutual inhibition of REM-on and REM-off populations	Human sleep: simulation of basic sleep patterns; sleep deprivation; and role of orexin (orexinergic neurons stabilize the wake-sleep cycle)
Quartet neural system model of sleep and wakefulness mechanisms Tamakawa et al.[127,173]	Neural system consisting of sleep-active preoptic or anterior hypothalamic neurons (N-R group); wake-active hypothalamic and brainstem neurons (WA group); brainstem neurons (REM group); and basal forebrain, hypothalamic, and brainstem neurons (W-R group) WA neurons have mutual inhibitory couplings with REM and N-R neurons. W-R neurons have mutual excitatory couplings with WA and REM neurons. REM neurons receive unidirectional inhibition from N-R neurons. N-R neurons are activated by sleep-promoting substances Reciprocal interaction between REM-on and REM-off populations	Human sleep: simulation of basic sleep-wakefulness rhythms Rodent sleep: model reproduced sleep and wakefulness patterns of rats; microinjection experiments; circadian modulation
Mathematical model of network dynamics governing mouse sleep-wake behavior Diniz Behn et al.[124,174]	Sleep-wake network composed of coupled relaxation oscillators Network includes wake-, sleep-, and REM sleep–promoting populations Incorporation of orexin signaling Reciprocal interaction between REM-on and REM-off populations	Rodent sleep: simulation of dynamics underlying state transitions
Modeling framework of the sleep-wake regulatory network in the brainstem and hypothalamus Diniz Behn & Booth[128,175-179]	Neurotransmitter-mediated interactions among brainstem and hypothalamic neuronal populations that participate in the transitions between wake, REM sleep, and NREM sleep Reciprocal interaction between REM-on and REM-off populations	Human sleep: simulation of basic sleep behavior and internal desynchronization Rodent sleep: simulation of basic sleep behavior; microinjection experiments; circadian modulation; temporal architecture of sleep patterns

CRF, Caudal reticular formation; LC, locus coeruleus; LDT/PPT, laterodorsal tegmentum/pedunculopontine tegmentum; MRF, midbrain reticular formation; POAH, pre-optic anterior hypothalamus; VLPO, ventrolateral preoptic area.

*Basic assumption in all models is that transitions between sleep or NREM sleep and wake are mediated by mutual inhibition between sleep or NREM sleep–promoting and wake-promoting neuronal populations. Inclusion of homeostatic and circadian components.

model; Table 36-1[97,98]). The model has been expanded[99] to account for discrepancies observed in chronic sleep restriction experiments.

Jewett and Kronauer (Table 36-1[100]) proposed interactive mathematical models of subjective alertness and cognitive throughput in humans. A homeostatic component (H) falls in a sigmoidal manner during waking and rises according to a saturating exponential function during sleep. The rise of H during sleep is determined by the circadian phase. H interacts with a circadian component (C[101]), accounting for the effect of light on the circadian pacemaker. The amplitude of C depends on the level of H. In addition, a sleep inertia component (W) is included. In contrast to the two- and three-process models, a nonlinear interaction is assumed. Whether the interaction is linear or nonlinear is still unresolved.[102,103] A statistical approach to test for nonlinear interactions was proposed[73] and is based on a comparison of model predictions and empiric data.

Experiments with chronic sleep restriction demonstrated that the homeostatic sleep drive is not associated with neurobehavioral performance.[104] The latter showed a cumulative impairment with a considerable degree of individual variability. To account for the slow recovery of neurobehavioral functions after sleep deprivation, the possibility of a second homeostatic process with a very long time constant was considered.[105] McCauley and colleagues[72,106] expanded the two-process model to a broader class of models that have a dynamic repertoire capturing waking neurobehavioral functions across a wide range of wake-sleep schedules. Models of sleep, fatigue, and performance based in large part on the two-process model are reported in a special issue of *Aviation, Space and Environmental Medicine*.[76] For a recent review, see Dawson et al.[107]

Because most of these models are based on average data, predictions of individual performance are limited.[108] A difficulty in addressing individual differences is due to the fact that individual parameters need first to be determined. Two approaches to predict individual performance in sleep deprivation experiments have been proposed.[109,110] Both approaches are based on the two-process model and allow a continuous parameter adaptation in an iterative process as new empiric data become available.

Although the qualitative version of the two-process model had originated from animal data,[2,81] the quantitative version was elaborated on the basis of findings from human studies. In the meantime, quantitative simulations of NREM sleep homeostasis were also performed in rats[111-114] and mice[115,116] (Figure 36-5, *B*). SWA of consecutive 4-second epochs in a 24-hour baseline, a 6-hour sleep deprivation, and 18-hour recovery period[115] served as the database for the simulation in mice. Like in the original human version of the model, process S was assumed to decrease exponentially in NREM and REM sleep and to increase according to a saturating exponential function in waking. Unlike in the human model, S increased also in REM sleep. After optimizing the initial value of S (IV) as well as its time constants (increase Ti; decrease Td), a close fit was obtained between the hourly mean values of SWA in NREM sleep and the prediction of process S (see Figure 36-5, *B*).

Modeling REM Sleep

Whereas many models have focused on NREM sleep homeostasis, on the interaction of NREM and REM sleep, and on the circadian oscillator, REM sleep regulation per se has been largely ignored. This is because of the complex and not yet clearly understood principles underlying REM sleep regulation (see earlier). The NREM-REM sleep cycle has been accounted for by the limit cycle reciprocal interaction model on the basis of interacting neuronal systems (for a elaborated version and for the activation-synthesis model, see Pace-Schott and Hobson[117]). Attempts were made to integrate various concepts into a combined model.[118,119] Saper and coworkers proposed in their flip-flop model a mode of interaction between different neurotransmitter systems that accounts for the sharp state transitions between waking and sleep[120] as well as between NREM and REM sleep.[121,122] Mathematical models incorporating these neurophysiologic aspects are summarized in Table 36-2.

Neurophysiologic Models

Neuronal models may be subdivided into neurophysiologic models based on detailed neuronal architecture (e.g., neurons, synapses) addressing thalamocortical interactions[123] and models that capture essential features based on the dynamics of neuronal populations (approximation of activity of a large ensemble of neurons; see Table 36-2).

Extended versions of the two-process model were derived from neurophysiologic data[124-129] (see Table 36-2). For a recent review and mathematical analysis of the models, see Booth and Diniz Behn.[130] Interactions of neuronal populations in the brainstem and hypothalamus were used to simulate transitions between sleep and waking or between NREM and REM sleep. In addition, a homeostatic and a circadian component was included. To a large degree, the slow dynamics (hours to days) of these models resemble the dynamics of the two-process model.[130,131] The degree of similarity between the two-process model and the Phillips-Robinson model[125] was recently explored.[131] The authors demonstrated that the slow dynamics of the Phillips-Robinson model could be explicitly related to the two-process model providing a neurophysiologic interpretation of the thresholds. This model[125] (see Table 36-2) was applied to simulate sleep fragmentation experiments,[132] differences in mammalian sleep patterns,[133] and subjective fatigue during sleep deprivation.[134] It was refined to simulate effects of caffeine[135] and spontaneous internal desynchronization through feedback of the sleep-wake cycle on the circadian component.[136,137]

For a better understanding of sleep-wake regulation, the development and analysis of models based on neurophysiologic mechanisms are essential. However, these models become more and more complex and comprise a large parameter space with only few parameters that can be determined empirically. Additionally, large numbers of simulations have to be performed to establish the system behavior. Therefore it is difficult to compare the performance of the different models. A further challenge is model validation. What are the relevant end points to investigate? The replication of qualitative patterns of behavioral state transitions is insufficient for parameter estimation of such highly complex models. Thus the application of sophisticated statistical approaches to determine model parameters is needed for further progress.[130]

A large-scale network[123] based on detailed neuronal architecture (65,400 neurons; 4,860,450 connections) has been proposed. It encompasses portions of two visual areas and associated thalamic and reticular thalamic nuclei addressing

the generation of slow oscillations (slower than 1 Hz). The areas are characterized by slow membrane potential fluctuations of cortical neurons, with depolarizing (up-state) and hyperpolarizing (down-state) components alternating with a frequency slower than 1 Hz. The model was used to simulate effects of neuromodulators or changes in synaptic strength[43] in slow waves. Furthermore, in a modified model encompassing three-layered motor cortex, effects of transcranial magnetic stimulation were investigated.[138]

CONCLUSIONS AND PERSPECTIVES

Ever since the first EEG recording by Hans Berger with a Siemens galvanometer, technical advances were instrumental for progress in sleep research. This has been evident in regard to the issues addressed in this chapter. The increasing ease of recording the sleep EEG from multiple derivations, combined with access to computer programs to analyze and visualize the data, has made it possible to explore the sleep process on a regional level. Differences became readily apparent between frontal and occipital derivations.[139,140] Frontal derivations have been shown to exhibit the largest response to changes in sleep pressure in terms of both SWA during sleep and theta activity during waking.[141-144] The topographic analysis of the EEG revealed that the frontal predominance of low-frequency activity is a common feature of NREM sleep, REM sleep, and waking and therefore represents a state-independent trait.[142,145] Thus basic features of the homeostatic process are present throughout the sleep-wake cycle, suggesting that sleep and waking could be seen as part of a continuum.[144]

Regional differences pose a challenge to modeling because differences in the characteristics of the homeostatic process (e.g., time constants, level of asymptote) may reflect different regulatory features of specific neuronal ensembles.[87,89]

The report of unihemispheric deep sleep in the dolphin[146,147] not only constituted evidence for a regional segregation of the sleep process but also raised the question of its functional significance. Hypotheses have been advanced implying that regional increases in neuronal activity and metabolic demand during wakefulness might result in selective changes in EEG synchronization of these neuronal populations during NREM sleep.[148-153]

The theory of a local, use-dependent increase of sleep intensity was tested by investigating whether a local activation of a particular brain region during wakefulness affects the EEG recorded from the same site during sleep. The first positive result consisted of the increase in low-frequency activity over the contralateral somatosensory cortex in the first hour of sleep following a vibratory stimulus to the contralateral hand.[154] This regional use-dependent effect was subsequently confirmed and extended.[155,156] Analogous findings were obtained in the rat and human, in which unilateral sensory or optic stimuli, respectively, during waking caused an interhemispheric shift in low-frequency power in the NREM sleep EEG.[157,158] Conversely, in a human study the selective understimulation of the cortical arm projection area during waking achieved by unilateral arm immobilization induced a reduction of power over the corresponding cortical region during subsequent sleep.[159]

The notion of sleep homeostasis, originally derived from the sleep-wake-dependent changes of the EEG slow waves, was recently expanded to the synaptic level. Tononi and Cirelli[151,152] proposed a synaptic homeostasis hypothesis postulating that synaptic strength is maintained over time by alternating phases of predominant potentiation during waking with phases of predominant depression during sleep. NREM sleep and the typical slow waves would subserve synaptic downscaling and thereby safeguard energy, space, and cellular supplies. The synaptic homeostasis hypothesis has the merit of relating the changes at the level of the EEG to well-known mechanisms at the synaptic level and thereby allowing specific predictions that can be not only simulated,[43] but also tested by electrophysiologic and neurochemical techniques in both humans and animals.[160-163] Vyazovskiy and Harris[153] proposed that periods of reduced synaptic input (so-called off periods or down states) serve prophylactic cellular maintenance. Krueger and collaborators[164-166] also underline the local aspect of sleep. Starting from an early theoretical paper,[149] they view sleep as an emergent property of cortical columns and propose a nonlinear mathematical model to account for the interactions between columns.[150]

In conclusion, models have proved useful for delineating regulating processes underlying such a complex and little-understood phenomenon as sleep and thereby offer a conceptual framework for analyzing existing and new data. The major models have already inspired a considerable number of experiments.

CLINICAL PEARL

The common experience that a good night's sleep dissipates fatigue and tiredness and regenerates energy points to a specific restorative function of sleep that cannot be achieved by merely resting. Sleep homeostasis denotes a basic principle of sleep regulation that can lead to a better understanding of sleep pathologies. Deficient sleep homeostasis might account for the altered sleep architecture in depression, and the transient normalization of sleep propensity can explain the antidepressant effect of sleep deprivation. The elucidation of sleep homeostasis at the cellular and molecular levels could open new avenues for the pharmacologic therapy of sleep disorders.

SUMMARY

Sleep homeostasis denotes a basic principle of sleep regulation. A sleep deficit elicits a compensatory increase in the intensity and duration of sleep, whereas excessive sleep reduces sleep propensity. It is as though sleep pressure is maintained within a range delimited by an upper and lower threshold. Sleep homeostasis is represented in the two-process model of sleep regulation by process S, which increases during waking and declines during sleep. The timing and propensity of sleep are modulated also by a circadian process. EEG SWA serves as an indicator of sleep homeostasis in NREM sleep. The level of SWA, a correlate of sleep intensity, is determined by the duration of prior sleep and waking. In the waking EEG, theta activity shows a rising trend with the progression of wakefulness and represents a marker of process S. However, in contrast to SWA, theta activity undergoes a marked circadian modulation. Advanced versions of the two-process model were applied to simulate the SWA pattern in a variety of experimental schedules. Other models addressing sleep regulation are derived from neurophysiologic data.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation and nano-tera.ch.

Selected Readings

Achermann P, Borbély AA. Low-frequency (<1 Hz) oscillations in the human sleep EEG. *Neuroscience* 1997;**81**:213–22.

Blake H, Gerard RW. Brain potentials during sleep. *Am J Physiol* 1937;**119**: 692–703.

Booth V, Diniz Behn CG. Physiologically-based modeling of sleep-wake regulatory networks. *Math Biosci* 2014;**250**:54–68.

Borbély AA. A two process model of sleep regulation. *Hum Neurobiol* 1982;**1**:195–204.

Borbély AA, Baumann F, Brandeis D, et al. Sleep deprivation: effect on sleep stages and EEG power density in man. *Electroencephalogr Clin Neurophysiol* 1981;**51**:483–93.

Daan S, Beersma DGM, Borbély AA. Timing of human sleep: recovery process gated by a circadian pacemaker. *Am J Physiol* 1984;**246**:R161–78.

Dijk DJ, Czeisler CA. Contribution of the circadian pacemaker and the sleep homeostat to sleep propensity, sleep structure, electroencephalographic slow waves, and sleep spindle activity in humans. *J Neurosci* 1995;**15**: 3526–38.

Mang GM, Franken P. Genetic dissection of sleep homeostasis. *Curr Top Behav Neurosci* 2015;**25**:25–63.

Olbrich E, Claussen JC, Achermann P. The multiple time scales of sleep dynamics as a challenge for modelling the sleeping brain. *Philos Trans A Math Phys Eng Sci* 2011;**369**:3884–901.

Phillips AJ, Robinson PA. A quantitative model of sleep-wake dynamics based on the physiology of the brainstem ascending arousal system. *J Biol Rhythms* 2007;**22**:167–79.

Ramakrishnan S, Lu W, Laxminarayan S, et al. Can a mathematical model predict an individual's trait-like response to both total and partial sleep loss? *J Sleep Res* 2015;**24**:262–9.

Skeldon AC, Dijk DJ, Derks G. Mathematical models for sleep-wake dynamics: comparison of the two-process model and a mutual inhibition neuronal model. *PLoS ONE* 2014;9:e103877.

Tononi G, Cirelli C. Sleep function and synaptic homeostasis. *Sleep Med Rev* 2006;**10**:49–62.

Van Dongen HPA, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep* 2003;**26**:117–26.

Vyazovskiy VV, Harris KD. Sleep and the single neuron: the role of global slow oscillations in individual cell rest. *Nat Rev Neurosci* 2013;**14**: 443–51.

A complete reference list can be found online at ExpertConsult.com.